

2023 IEEE 41st International Conference on Computer Design (ICCD 2023)

**Washington, DC, USA
6-8 November 2023**



**IEEE Catalog Number: CFP23ICD-POD
ISBN: 979-8-3503-4292-5**

**Copyright © 2023 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP23ICD-POD
ISBN (Print-On-Demand):	979-8-3503-4292-5
ISBN (Online):	979-8-3503-4291-8
ISSN:	1063-6404

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2023 IEEE 41st International Conference on Computer Design (ICCD) ICCD 2023

Table of Contents

Message from the General Chairs	xix
Message from the Program Chairs	xx
Organizing Committee	xxii
Program Committee	xxiii
Keynotes	xxviii
Sponsors	xxxii

Session 1A: Verification & Security

Leveraging Firmware Reverse Engineering for Stealthy Sensor Attacks via Binary Modification	1
<i>Sutej Kulkarni (University of California Davis, USA), Ryan Tsang (University of California Davis, USA), Asmita Asmita (University of California Davis, USA), Houman Homayoun (University of California Davis, USA), and Soheil Salehi (University of Arizona, USA)</i>	
A Compressed and Accurate Sparse Deep Learning Based Workload-Aware Timing Error Model	9
<i>Styliani Tompazi (Queen's University Belfast, UK) and Georgios Karakonstantis (Queen's University Belfast, UK)</i>	
Transcend Adversarial Examples: Diversified Adversarial Attacks to Test Deep Learning Model	13
<i>Wei Kong (National Key Laboratory on Science and Technology of Information System Security)</i>	
REMU: Enabling Cost-Effective Checkpointing and Deterministic Replay in FPGA-based Emulation	21
<i>Yuxiao Chen (State Key Lab of Processors, Institute of Computing Technology, CAS, China; University of Chinese Academy of Sciences, China), Yisong Chang (State Key Lab of Processors, Institute of Computing Technology, CAS, China; University of Chinese Academy of Sciences, China), Ke Zhang (State Key Lab of Processors, Institute of Computing Technology, CAS, China; University of Chinese Academy of Sciences, China), Mingyu Chen (State Key Lab of Processors, Institute of Computing Technology, CAS, China; University of Chinese Academy of Sciences, China), and Yungang Bao (State Key Lab of Processors, Institute of Computing Technology, CAS, China; University of Chinese Academy of Sciences, China)</i>	

Model Checking TileLink Cache Coherence Protocols By Murphi	30
<i>Zimin Li (Institute of Software, Chinese Academy of Sciences, China; Hangzhou Institute for Advanced Study, China), Yongjian Li (Institute of Software, Chinese Academy of Sciences, China), Kaifan Wang (Institute of Computer Technology, Chinese Academy of Sciences, China; Beijing Institute of Open Source Chip, China), Kun Ma (Institute of Software, Chinese Academy of Sciences, China), and Shizhen Yu (Institute of Software, Chinese Academy of Sciences, China)</i>	

Session 1B: Logic and Circuit Design

MNHOKA - PPA Efficient M-Term Non-Homogeneous Hybrid Overlap-Free Karatsuba Multiplier for GF (2 ⁿ) Polynomial Multiplier	38
<i>Gogireddy Ravi Kiran Reddy (International Institute of Information Technology Bangalore, India), Sanampudi Gopala Krishna Reddy (International Institute of Information Technology Bangalore, India), Vasanthi D R (International Institute of Information Technology Bangalore, India), and Madhav Rao (International Institute of Information Technology Bangalore, India)</i>	
ApproxCNN: Evaluation Of CNN With Approximated Layers Using In-Exact Multipliers	46
<i>Bindu G Gowda (IIIT-Bangalore, India), Raghava S N (IIIT-Bangalore, India), Prashanth H C (IIIT-Bangalore, India), Pratyush Nandi (IIIT-Bangalore, India), and Madhav Rao (IIIT-Bangalore, India)</i>	
ACET: An Adaptive Clock Scheme Exploiting Comprehensive Timing Slack for Reconfigurable Processors	54
<i>Shuya Ji (Shanghai Jiao Tong University, China), Weidong Yang (Shanghai Jiao Tong University, China), Jianfei Jiang (Shanghai Jiao Tong University, China), Naifeng Jing (Shanghai Jiao Tong University, China), Weiguang Sheng (Shanghai Jiao Tong University, China), Ang Li (Shanghai Jiao Tong University, China), and Qin Wang (Shanghai Jiao Tong University, China)</i>	
SFDoP: A Scalable Fused BFloat16 Dot-Product Architecture for DNN	62
<i>Jing Zhang (National University of Defense Technology, China), Hongbing Tan (National University of Defense Technology, China), and Libo Huang (National University of Defense Technology, China)</i>	
ImprLM: An Improved Logarithmic Multiplier Design Approach via Iterative Linear-Compensation and Modified Dynamic Segment	66
<i>Shangshang Yao (National University of Defense Technology, China) and Li Shen (National University of Defense Technology, China)</i>	

Session 2A: Brain-Inspired Circuits

MindCrypt: The Brain as a Random Number Generator for SoC-Based Brain-Computer Interfaces ..	70
<i>Guy Eichler (Columbia University, USA), Biruk Seyoum (Columbia University, USA), Kuan-Lin Chiu (Columbia University, USA), and Luca P. Carloni (Columbia University, USA)</i>	

BrainTTA: A 28.6 TOPS/W Compiler Programmable Transport-Triggered NN SoC	78
<i>Maarten J. Molendijk (Eindhoven University of Technology, the Netherlands), Floran A.M. de Putter (Eindhoven University of Technology, the Netherlands), Manil Dev Gomony (Eindhoven University of Technology, the Netherlands), Pekka Jääskeläinen (Tampere University, Finland), and Henk Corporaal (Eindhoven University of Technology, the Netherlands)</i>	

Session 2B: Quantum Computing

HiSEP-Q: A Highly Scalable and Efficient Quantum Control Processor for Superconducting Qubits	86
<i>Xiaorang Guo (Technical University of Munich), Kun Qin (Technical University of Munich), and Martin Schulz (Technical University of Munich; Leibniz Supercomputing Centre, Germany)</i>	
Enhancing Virtual Distillation with Circuit Cutting for Quantum Error Mitigation	94
<i>Peiyi Li (NC State University, USA), Ji Liu (Argonne National Laboratory, USA), Hrushikesh Pramod Patil (NC State University, USA), Paul Hovland (Argonne National Laboratory, USA), and Huiyang Zhou (NC State University, USA)</i>	

Session 3A: SRAM & NVM

ICON: An IR Drop Compensation Method at OU Granularity with Low Overhead for eNVM-based Accelerators	102
<i>Jinpeng Liu (Huazhong University of Science and Technology, China), Wei Tong (Huazhong University of Science and Technology, China), Bing Wu (Huazhong University of Science and Technology, China), Huan Cheng (Huazhong University of Science and Technology, China), and Chengning Wang (Huazhong University of Science and Technology, China)</i>	
Resonant Compute-In-Memory (rCIM) 10T SRAM Macro for Boolean Logic	110
<i>Dhandeep Challagundla (University of Maryland, Baltimore County), Ignatius Bezzam (Rezonent Inc.), Biprangshu Saha (Plus PPA Nanochip Technologies), and Riadul Islam (University of Maryland, Baltimore County)</i>	
Small Footprint 6T-SRAM Design with MIV-Transistor Utilization in M3D-IC Technology	118
<i>Madhava Sarma Vemuri (North Dakota State University, North Dakota) and Umamaheswara Rao Tida (North Dakota State University, North Dakota)</i>	

Session 3B: Cache Memory

Offline and Online Algorithms for Cache Allocation with Monte Carlo Tree Search and a Learned Model	126
<i>Yibin Gu (Huazhong University of Science and Technology, China), Hua Wang (Huazhong University of Science and Technology, China), Man Luo (Huazhong University of Science and Technology, China), Jingyu Tang (Huazhong University of Science and Technology, China), and Ke Zhou (Huazhong University of Science and Technology, China)</i>	

Morpheus: An Adaptive DRAM Cache with Online Granularity Adjustment for Disaggregated Memory	134
<i>Xu Zhang (Institute of Computing Technology, China; University of Chinese Academy of Sciences (UCAS), China), Tianyue Lu (Institute of Computing Technology, China; University of Chinese Academy of Sciences (UCAS), China), Yisong Chang (Institute of Computing Technology, China; University of Chinese Academy of Sciences (UCAS), China; Zhongguancun Laboratory, China), Ke Zhang (Institute of Computing Technology, China; University of Chinese Academy of Sciences (UCAS), China), and Mingyu Chen (Institute of Computing Technology, China; University of Chinese Academy of Sciences (UCAS), China; Zhongguancun Laboratory, China)</i>	
Locality-Aware Speculative Cache for Fast Partial Updates in Erasure-Coded Cloud Clusters	142
<i>Hai Zhou (Huazhong University of Science and Technology, China), Yuchong Hu (Huazhong University of Science and Technology, China), Dan Feng (Huazhong University of Science and Technology, China), Wei Wang (Hikvision, China), and Huadong Huang (Hikvision, China)</i>	

Session 4A: Accelerators

A Cost-Efficient Failure-Tolerant Scheme for Distributed DNN Training	150
<i>Menglei Chen (Huazhong University of Science and Technology, China), Yu Hua (Huazhong University of Science and Technology, China), Rong Bai (Huazhong University of Science and Technology, China), and Jianming Huang (Huazhong University of Science and Technology, China)</i>	
RealArch: A Real-Time Scheduler for Mapping Multi-Tenant DNNs on Multi-Core Accelerators ...	158
<i>Xuhang Wang (Shanghai Jiao Tong University, China), Zhuoran Song (Shanghai Jiao Tong University, China), and Xiaoyao Liang (Shanghai Jiao Tong University, China)</i>	
Polyform: A Versatile Architecture for Multi-DNN Execution via Spatial and Temporal Acceleration	166
<i>Lingxiang Yin (University of Central Florida, USA), Amir Ghazizadeh (University of Central Florida, USA), Shilin Tian (University of Central Florida, USA), Ahmed Louri (George Washington University, USA), and Hao Zheng (University of Central Florida, USA)</i>	
STAG: Enabling Low Latency and Low Staleness of GNN-based Services with Dynamic Graphs ...	170
<i>Jiawen Wang (Shanghai Jiao Tong University, China), Quan Chen (Shanghai Jiao Tong University, China), Deze Zeng (China University of Geosciences, China), Zhuo Song (Alibaba Cloud, China), Chen Chen (Shanghai Jiao Tong University, China), and Minyi Guo (Shanghai Jiao Tong University, China)</i>	

Session 4B: Persistent Memory

Accelerating Persistent Hash Indexes via Reducing Negative Searches	174
<i>Renzhi Xiao (Huazhong University of Science and Technology, China), Hong Jiang (University of Texas at Arlington, USA), Dan Feng (Huazhong University of Science and Technology, China), Yuchong Hu (Huazhong University of Science and Technology, China), Wei Tong (Huazhong University of Science and Technology, China), Kang Liu (Huazhong University of Science and Technology, China), Yucheng Zhang (Huazhong University of Science and Technology, China), Xueliang Wei (Huazhong University of Science and Technology, China), and Zhengtao Li (Huazhong University of Science and Technology, China)</i>	
PMA: A Persistent Memory Allocator with High Efficiency and Crash Consistency Guarantee	182
<i>Xiangyu Xiang (Huazhong University of Science and Technology, China), Yu Hua (Huazhong University of Science and Technology, China), and Hao Xu (Huazhong University of Science and Technology, China)</i>	
Prediction-Guided Metadata Backup for Improving Lifetime on Flash-based Swap	190
<i>Taejoon Song (LG Electronics, South Korea), June-Hyung Kim (LG Electronics, South Korea), Myeongseon Kim (LG Electronics, South Korea), and Youngjin Kim (LG Electronics, South Korea)</i>	
RWORT: A Read and Write Optimized Radix Tree for Persistent Memory	194
<i>Jinlei Hu (Huazhong University of Science and Technology, China), Zijie Wei (Huazhong University of Science and Technology, China), Jianxi Chen (Huazhong University of Science and Technology, China), and Dan Feng (Huazhong University of Science and Technology, China)</i>	

Session 5A: Storage

An Effective and Balanced Storage Extension Approach for Sharding Blockchain Systems	198
<i>Tingyu Fan (Tianjin University, China), Xiulong Liu (Tianjin University, China), Baochao Chen (Tianjin University, China), and Wenyu Qu (Tianjin University, China)</i>	
BlzFS: Crash Consistent Log-Structured File System Based on Byte-Loggable Zone for ZNS SSD....	206
<i>Wenjie Qi (Huazhong University of Science and Technology, China), Zhipeng Tan (Huazhong University of Science and Technology, China), Ziyue Zhang (Huazhong University of Science and Technology, China), Jing Zhang (Huazhong University of Science and Technology, China), Chao Yu (Guangdong Oppo Mobile Telecommunications Corp., Ltd, China), Ying Yuan (Huazhong University of Science and Technology, China), and Shikai Tan (Huazhong University of Science and Technology, China)</i>	
K8sES: Optimizing Kubernetes with Enhanced Storage Service-Level Objectives	214
<i>Hao Wen (University of Minnesota, Twin Cities), Zhichao Cao (Arizona State University), Bingzhe Li (University of Texas at Dallas), David H.C. Du (University of Minnesota, Twin Cities), Ayman Abouelwafa (Hewlett Packard Enterprise), Doug Voigt (Hewlett Packard Enterprise), Shiyong Liu (University of Minnesota, Twin Cities), Jim Diehl (University of Minnesota, Twin Cities), and Fenggang Wu (University of Minnesota, Twin Cities)</i>	

Session 5B: Memory Systems

CostFM: A High Cost-Performance Fingerprint Management Mechanism for Shared SSDs	223
<i>Hao Liu (Huazhong University of Science and Technology, China), Mengting Lu (Huazhong University of Science and Technology, China; Alibaba Cloud Intelligence Group, China), Fang Wang (Huazhong University of Science and Technology, China), and Wenpeng He (Huazhong University of Science and Technology, China)</i>	
Enabling Encrypted Delta Compression for Outsourced Storage Systems via Preserving Similarity	231
<i>Chuang Gan (Huazhong University of Science & Technology, China), Yuchong Hu (Huazhong University of Science & Technology, China), Leyan Zhao (Huazhong University of Science & Technology, China), Xin Zhao (Huazhong University of Science & Technology, China), Pengyu Gong (Huazhong University of Science & Technology, China), Wenhao Zhang (Huazhong University of Science & Technology, China), Lin Wang (Huazhong University of Science & Technology, China), and Dan Feng (Huazhong University of Science & Technology, China)</i>	
FlashDAM: Flexible I/O Throttling for the User Experience of Mobile Systems	239
<i>Changlong Li (East China Normal University, China), Chao Wang (University of Science and Technology of China, China), Xuehai Zhou (University of Science and Technology of China, China), and Edwin H.-M. Sha (East China Normal University, China)</i>	
HyperMetric: Robust Hyperdimensional Computing on Error-Prone Memories using Metric Learning	243
<i>Weihong Xu (University of California San Diego, USA), Viji Swaminathan (University of California San Diego, USA), Sumukh Pinge (University of California San Diego, USA), Sean Fuhrman (University of California San Diego, USA), and Tajana Rosing (University of California San Diego, USA)</i>	

Session 6A: GPU & Graph

KeSCo: Compiler-based Kernel Scheduling for Multi-Task GPU Applications	247
<i>ZeJia Lin (Sun Yat-sen University, China), Zewei Mo (University of Pittsburgh, USA), Xuanteng Huang (Sun Yat-sen University, China), Xianwei Zhang (Sun Yat-sen University, China), and Yutong Lu (Sun Yat-sen University, China)</i>	
Beyond Compression Ratio: A Throughput Analysis of Memory Compression Techniques for GPUs ...	255
<i>Manuel Renz (Technical University of Hamburg, Germany) and Sohan Lal (Technical University of Hamburg, Germany)</i>	

PANG: A Pattern-Aware GCN Accelerator for Universal Graphs	263
<i>Yibo Du (CICS, Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Ying Wang (CICS, Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Shengwen Liang (SKLP, Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Huawei Li (SKLP, Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Peng Cheng Laboratory, China), Xiaowei Li (SKLP, Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), and Yinhe Han (CICS, Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Zhejiang Lab, China)</i>	

Session 6B: File Systems

HyF2FS: A Filesystem to Fully Exploit the Parallelism of Hybrid Storage	267
<i>Jintong Zhang (Huazhong University of Science and Technology, China), Jianxi Chen (Huazhong University of Science and Technology, China), Kezheng Liu (Huazhong University of Science and Technology, China), Yongkang Zhuo (Huazhong University of Science and Technology, China), and Panfei Yuan (Huazhong University of Science and Technology, China)</i>	
SMRTS: A Performance and Cost-Effectiveness Optimized SSD-SMR Tiered File System with Data Deduplication	275
<i>Zhichao Cao (Arizona State University), Hao Wen (University of Minnesota, Twin Cities), Fenggang Wu (University of Minnesota, Twin Cities), and David H.C. Du (University of Minnesota, Twin Cities)</i>	
Low-Latency and Scalable Full-Path Indexing Metadata Service for Distributed File Systems	283
<i>Chao Dong (Huazhong University of Science and Technology, China), Fang Wang (Huazhong University of Science and Technology, China), Yuxin Yang (Huazhong University of Science and Technology, China), Mengya Lei (Huazhong University of Science and Technology, China), Jianshun Zhang (Huazhong University of Science and Technology, China), and Dan Feng (Huazhong University of Science and Technology, China)</i>	

Session 7A: SSDs

FlexZNS: Building High-Performance ZNS SSDs with Size-Flexible and Parity-Protected Zones	291
<i>Yu Wang (Huazhong University of Science and Technology, China), You Zhou (Huazhong University of Science and Technology, China), Zhonghai Lu (KTH Royal Institute of Technology, Sweden), Xiaoyi Zhang (Alibaba Group, China), Kun Wang (Alibaba Group, China), Feng Zhu (Alibaba Group, China), Shu Li (Alibaba Group, China), Changsheng Xie (Huazhong University of Science and Technology, China), and Fei Wu (Huazhong University of Science and Technology, China)</i>	

LifetimeKV: Narrowing the Lifetime Gap of SSTs in LSMT-based KV Stores for ZNS SSDs	300
<i>Biyong Liu (Huazhong University of Science & Technology, China), Yuan Xia (Huazhong University of Science & Technology, China), Xueliang Wei (Huazhong University of Science & Technology, China), and Wei Tong (Huazhong University of Science & Technology, China)</i>	
Persimmon: An Append-Only ZNS-First Filesystem	308
<i>Devashish R. Purandare (University of California, USA), Sam Schmidt (University of California, USA), and Ethan L. Miller (University of California, USA; Pure Storage, USA)</i>	
Turn Waste Into Wealth: Alleviating Read/Write Interference in ZNS SSDs	316
<i>Weilin Zhu (Huazhong University of Science and Technology, China) and Wei Tong (Huazhong University of Science and Technology, China)</i>	

Session 7B: Logic Synthesis

GPT-LS: Generative Pre-Trained Transformer with Offline Reinforcement Learning for Logic Synthesis	320
<i>Chenyang Lv (Shanghai Jiao Tong University, China), Ziling Wei (Shanghai Jiao Tong University, China), Weikang Qian (Shanghai Jiao Tong University, China), Junjie Ye (Huawei Noah's Ark Lab, China), Chang Feng (Huawei Noah's Ark Lab, China), and Zhezhi He (Shanghai Jiao Tong University, China)</i>	
Delay-Driven Physically-Aware Logic Synthesis with Informed Search	327
<i>Linyu Zhu (University of Michigan, China; Shanghai Jiao Tong University, China) and Xinfei Guo (University of Michigan, China; Shanghai Jiao Tong University, China)</i>	
Adaptive Reconvergence-Driven AIG Rewriting via Strategy Learning	336
<i>Liwei Ni (Institute of Computing Technology, Chinese Academy of Sciences, China; Peng Cheng Laboratory, China; University of Chinese Academy of Sciences, China), Zonglin Yang (Shenzhen University, China), Jiayi Zhang (Peking University, China), Junfeng Liu (Beihang University, China), Huawei Li (Institute of Computing Technology, Chinese Academy of Sciences, China; Peng Cheng Laboratory, China; University of Chinese Academy of Sciences, China), Biwei Xie (Institute of Computing Technology, Chinese Academy of Sciences, China; Peng Cheng Laboratory, China; University of Chinese Academy of Sciences, China), and Xingquan Li (Peng Cheng Laboratory, China)</i>	
AiMap: Learning to Improve Technology Mapping for ASICs via Delay Prediction	344
<i>Junfeng Liu (Beihang University, China), Liwei Ni (Peng Cheng National Laboratory, China), Xingquan Li (Peng Cheng National Laboratory, China), Min Zhou (Huawei Noah's Ark Lab, China), Lei Chen (Huawei Noah's Ark Lab, China), Xing Li (Huawei Noah's Ark Lab, China), Qinghua Zhao (Beihang University, China), and Shuai Ma (Beihang University, China)</i>	

Session 8A: GPU II

FlexGM: An Adaptive Runtime System to Accelerate Graph Matching Networks on GPUs	348
<i>Yue Dai (University of Pittsburgh, PA), Xulong Tang (University of Pittsburgh, PA), and Youtao Zhang (University of Pittsburgh, PA)</i>	
NTTFusion: Efficient Number Theoretic Transform Acceleration on GPUs	357
<i>Zhiwei Wang (State Key Laboratory of Information Security, Institute of Information Engineering, CAS, China; University of Chinese Academy of Sciences, China), Peinan Li (State Key Laboratory of Information Security, Institute of Information Engineering, CAS, China), Rui Hou (State Key Laboratory of Information Security, Institute of Information Engineering, CAS, China), and Dan Meng (State Key Laboratory of Information Security, Institute of Information Engineering, CAS, China)</i>	
MixRec: Orchestrating Concurrent Recommendation Model Training on CPU-GPU Platform	366
<i>Jiazhi Jiang (Sun Yat-sen University, China), Rui Tian (Sun Yat-sen University, China), Jiangsu Du (Sun Yat-sen University, China), Dan Huang (Sun Yat-sen University, China), and Yutong Lu (Sun Yat-sen University, China)</i>	

Session 8B: Accelerators

HyAcc: A Hybrid CAM-MAC RRAM-based Accelerator for Recommendation Model	375
<i>Xuan Zhang (Shanghai Jiao Tong University, China), Zhuoran Song (Shanghai Jiao Tong University, China), Xing Li (Shanghai Jiao Tong University, China), Zhezhi He (Shanghai Jiao Tong University, China), Li Jiang (Shanghai Jiao Tong University, China), Naifeng Jing (Shanghai Jiao Tong University, China), and Xiaoyao Liang (Shanghai Jiao Tong University, China)</i>	
ViTframe: Vision Transformer Acceleration via Informative Frame Selection for Video Recognition	383
<i>Chunyu Qi (Shanghai Jiao Tong University, China), Zilong Li (Shanghai Jiao Tong University, China), Zhuoran Song (Shanghai Jiao Tong University, China), and Xiaoyao Liang (Shanghai Jiao Tong University, China)</i>	
ACCO: Automated Causal CNN Scheduling Optimizer for Real-Time Edge Accelerators	391
<i>Jun Yin (ESAT-MICAS KU Leuven), Linyan Mei (ESAT-MICAS KU Leuven), Andre Guntoro (Robert Bosch GmbH), and Marian Verhelst (ESAT-MICAS KU Leuven)</i>	

Session 9A: Co-Design

Re-Compact: Structured Pruning and SpMM Kernel Co-Design for Accelerating DNNs on GPUs ..	399
<i>Yuling Zhang (Chongqing University, China), Ao Ren (Chongqing University, China), Xianzhang Chen (Chongqing University, China), Qiu Lin (Chongqing University, China), Yujuan Tan (Chongqing University, China), and Duo Liu (Chongqing University, China)</i>	

FM-P2L: An Algorithm Hardware Co-Design of Fixed-Point MSBs with Power-of-2 LSBs in CNN Accelerators	407
<i>Jun-Shen Wu (National Tsing Hua University, Taiwan) and Ren-Shuo Liu (National Tsing Hua University, Taiwan)</i>	
Hardware-Software Co-Design for Content-Based Sparse Attention	415
<i>Rui Tang (University of Science and Technology of China, China; Institute of Computing Technology, Chinese Academy of Sciences, China), Xiaoyu Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Rui Liu (Institute of Computing Technology, Chinese Academy of Sciences, China; Xiangtan University, China), Zhejiang Luo (Institute of Computing Technology, Chinese Academy of Sciences, China; Xiangtan University, China), Xiaoming Chen (Institute of Computing Technology, Chinese Academy of Sciences, China), and Yinhe Han (Institute of Computing Technology, Chinese Academy of Sciences, China)</i>	
Towards Quantized Stochastic Computing by Leveraging Reduced Precision Binary Numbers through Bit Truncation	419
<i>Donghui Lee (Kyungpook National University, Republic of Korea) and Yongtae Kim (Kyungpook National University, Republic of Korea)</i>	

Session 9B: Dataflow & Reinforcement Learning

Conveyor: Towards Asynchronous Dataflow in Systolic Array to Exploit Unstructured Sparsity....	423
<i>Seongwook Kim (Sungkyunkwan University, Korea), Gwangeun Byeon (Sungkyunkwan University, Korea), Sihyung Kim (Sungkyunkwan University, Korea), Hyungjin Kim (Sungkyunkwan University, Korea), and Seokin Hong (Sungkyunkwan University, Korea)</i>	
DFGC: DFG-Aware NoC Control based on Time Stamp Prediction for Dataflow Architecture	432
<i>Tianyu Liu (Institute of Computing Technology, China; University of Chinese Academy of Sciences, China), Wenming Li (Institute of Computing Technology, China; University of Chinese Academy of Sciences, China), and Zhihua Fan (Institute of Computing Technology, China; University of Chinese Academy of Sciences, China)</i>	
Alleviating Transfer Latency in DataFlow Accelerator for DSP Applications	440
<i>Haibin Wu (Institute of Computing Technology, China; UCAS, China), Wenming Li (Institute of Computing Technology, China; UCAS, China), Zhihua Fan (Institute of Computing Technology, China; UCAS, China), Zhen Wang (Institute of Computing Technology, China; UCAS, China), Tianyu Liu (Institute of Computing Technology, China; UCAS, China), Junying Huang (Institute of Computing Technology, China; UCAS, China), Shengzhong Tang (Institute of Computing Technology, China; UCAS, China), Yanhuan Liu (Institute of Computing Technology, China; UCAS, China), Kunming Zhang (Institute of Computing Technology, China; UCAS, China), Xiaochun Ye (Institute of Computing Technology, China; UCAS, China), and Dongrui Fan (Institute of Computing Technology, China; UCAS, China)</i>	

FLASH-RL: Federated Learning Addressing System and Static Heterogeneity using Reinforcement Learning	444
<i>Sofiane Bouaziz (École nationale Supérieure d'Informatique, Algeria), Hadjer Benmeziane (Univ. Polytechnique Hauts-de-France, France), Youcef Imine (Univ. Polytechnique Hauts-de-France, France), Leila Hamdad (École nationale Supérieure d'Informatique, Algeria), Smail Niar (Univ. Polytechnique Hauts-de-France, France), and Hamza Ouarnoughi (Univ. Polytechnique Hauts-de-France, France)</i>	

Session 10A: Matrix Multiplication & Sparsity

PrSpMV: An Efficient Predictable Kernel for SpMV	448
<i>Gelin Fu (Xi'an Jiaotong University, China), Tian Xia (Xi'an Jiaotong University, China), Shaoru Qu (Xi'an Jiaotong University, China), Zhongpei Luo (Xi'an Jiaotong University, China), Shuyu Li (Xi'an Jiaotong University, China), Pengyu Cheng (Xi'an Jiaotong University, China), Runfan Guo (Xi'an Jiaotong University, China), Yitong Ding (Xi'an Jiaotong University, China), and Pengju Ren (Xi'an Jiaotong University, China)</i>	
Releasing the Potential of Tensor Core for Unstructured SpMM using Tiled-CSR Format	457
<i>Zeyu Xue (National University of Defense Technology, China), Mei Wen (National University of Defense Technology, China), Zhaoyun Chen (National University of Defense Technology, China), Yang Shi (National University of Defense Technology, China), Minjin Tang (National University of Defense Technology, China), Jianchao Yang (National University of Defense Technology, China), and Zhongdi Luo (National University of Defense Technology, China)</i>	
Tailoring CUTLASS GEMM using Supervised Learning	465
<i>Yongseung Yu (Hanyang University, Korea), Donghyun Son (Hanyang University, Korea), Younghyun Lee (Hanyang University, Korea), Sunghyun Park (University of Michigan, USA), Giha Ryu (ATTO Research, Korea), Myeongjin Cho (SAPEON Korea Inc., Korea), Jiwon Seo (Hanyang University, Korea), and Yongjun Park (Yonsei University, Korea)</i>	

Session 10B: Fault Tolerance & Resilience

Revitalizing Buffered I/O: Optimizing Page Reclaim and I/O Throttling	475
<i>Jongseok Kim (Sungkyunkwan University, Republic of Korea), Chanu Yu (Sungkyunkwan University, Republic of Korea), and Euseong Seo (Sungkyunkwan University, Republic of Korea)</i>	
ResCheck: Resilient Checkpointing for Energy Harvesting Systems	483
<i>Keni Qiu (Capital Normal University, China), Chuting Xu (Capital Normal University, China), Kunyu Zhou (Capital Normal University, China), and Dehui Qiu (Capital Normal University, China)</i>	
Heart: A Scalable, High-Performance ART for Persistent Memory	487
<i>Liangxu Nie (Shanghai Jiao Tong University, China), Shengan Zheng (Shanghai Jiao Tong University, China), Bowen Zhang (Shanghai Jiao Tong University, China), Jinyan Xu (Shanghai Jiao Tong University, China), and Linpeng Huang (Shanghai Jiao Tong University, China)</i>	

DCR: Decomposition-Aware Column Re-Mapping for Stuck-At-Fault Tolerance in ReRAM Arrays 491

Hyeonsu Bang (Sungkyunkwan University, South Korea), Kang Eun Jeon (Sungkyunkwan University, South Korea), Johnny Rhe (Sungkyunkwan University, South Korea), and Jong Hwan Ko (Sungkyunkwan University, South Korea)

Snapshot: Fast, Userspace Crash Consistency for CXL and PM Using Msync 495

Suyash Mahar (UC San Diego), Mingyao Shen (UC San Diego), Terence Kelly (Independent), and Steven Swanson (UC San Diego)

Session 11A: Processing-In-Memory

GIM: Versatile GNN Acceleration with Reconfigurable Processing-in-Memory 499

Chen Nie (Shanghai Jiao Tong University, China), Guoyang Chen (Alibaba Cloud Infrastructure, USA), Weifeng Zhang (Alibaba Cloud Infrastructure, USA), and Zhezhi He (Shanghai Jiao Tong University, China)

PSQ: An Automatic Search Framework for Data-Free Quantization on PIM-based Architecture 507

Fangxin Liu (Shanghai Jiao Tong University, China; Shanghai Qi Zhi Institute, China), Ning Yang (Shanghai Jiao Tong University, China; Shanghai Qi Zhi Institute, China), and Li Jiang (Shanghai Jiao Tong University, China; Shanghai Qi Zhi Institute, China)

Exploiting and Enhancing Computation Latency Variability for High-Performance Time-Domain Computing-in-Memory Neural Network Accelerators 515

Chia-Chun Wang (National Tsing Hua University, Taiwan), Yun-Chen Lo (National Tsing Hua University, Taiwan), Jun-Shen Wu (National Tsing Hua University, Taiwan), Yu-Chih Tsai (National Tsing Hua University, Taiwan), Chia-Cheng Chang (National Tsing Hua University, Taiwan), Tsen-Wei Hsu (National Tsing Hua University, Taiwan), Min-Wei Chu (National Tsing Hua University, Taiwan), Chuan-Yao Lai (National Tsing Hua University, Taiwan), and Ren-Shuo Liu (National Tsing Hua University, Taiwan)

Input-Aware Flow-Based In-Memory Computing 523

Suraj Singireddy (University of Texas at San Antonio, USA), Muhammad Rashedul Haq Rashed (University of Central Florida, USA), Sven Thijssen (University of Central Florida, USA), Rickard Ewetz (University of Central Florida, USA), and Sumit K. Jha (Florida International University, USA)

BICEP: Exploiting Bitline Inversion for Efficient Operation-Unit-Based Compute-in-Memory Architecture: No Retraining Needed! 531

Yun-Chen Lo (National Tsing Hua University, Taiwan), Chia-Chun Wang (National Tsing Hua University, Taiwan), and Ren-Shuo Liu (National Tsing Hua University, Taiwan)

Cerasure: Fast Acceleration Strategies For XOR-Based Erasure Codes 535

Tianyang Niu (University of Science and Technology of China), Min Iyu (University of Science and Technology of China), Wei Wang (University of Science and Technology of China), Qiliang Li (University of Science and Technology of China), and Yinlong Xu (University of Science and Technology of China)

Session 11B: Electronic Design Automation

A DSP Shared is a DSP Earned: HLS Task-Level Multi-Pumping for High-Performance Low-Resource Designs	551
<i>Giovanni Brignone (Politecnico di Torino, Italy), Mihai T. Lazarescu (Politecnico di Torino, Italy), and Luciano Lavagno (Politecnico di Torino, Italy)</i>	
Efficient RISC-V-on-x64 Floating Point Simulation	558
<i>Niko Zurstraßen (RWTH Aachen University), Nils Bosbach (RWTH Aachen University), Jan Moritz Joseph (RWTH Aachen University), Lukas Jünger (MachineWare GmbH), Jan Henrik Weinstock (MachineWare GmbH), and Rainer Leupers (RWTH Aachen University)</i>	
HF-LDPC: HLS-Friendly QC-LDPC FPGA Decoder with High Throughput and Flexibility	566
<i>Yifan Zhang (Huazhong University of Science and Technology, China), Qiang Cao (Huazhong University of Science and Technology, China), Shaohua Wang (Huazhong University of Science and Technology, China), Jie Yao (Huazhong University of Science and Technology, China), and Hong Jiang (University of Texas at Arlington)</i>	
GNNHLS: Evaluating Graph Neural Network Inference via High-Level Synthesis	574
<i>Chenfeng Zhao (Washington University in St. Louis), Zehao Dong (Washington University in St. Louis), Yixin Chen (Washington University in St. Louis), Xuan Zhang (Washington University in St. Louis), and Roger Chamberlain (Washington University in St. Louis)</i>	

Session 12A: Test & Verification

Architectural Contracts for Safe Speculation	578
<i>Franz A. Fuchs (University of Cambridge, UK), Jonathan Woodruff (University of Cambridge, UK), Peter Rugg (University of Cambridge, UK), Marno van der Maas (University of Cambridge, UK), Alexandre Joannou (University of Cambridge, UK), Alexander Richardson (University of Cambridge, UK), Jessica Clarke (University of Cambridge, UK), Nathaniel Wesley Filardo (Microsoft Research Ltd, UK), Brooks Davis (SRI International, USA), John Baldwin (Ararat River Consulting, USA), Peter G. Neumann (SRI International, USA), Simon W. Moore (University of Cambridge, UK), and Robert N. M. Watson (University of Cambridge, UK)</i>	

Execute on Clear (EoC): Enhancing Security for Unsafe Speculative Instructions by Precise Identification and Safe Execution	587
<i>Xiaoni Meng (University of Chinese Academy of Sciences, China; Institute of Software, CAS, China), Qiusong Yang (Institute of Software, CAS, China), Yiwei Ci (Institute of Software, CAS, China), Pei Zhao (Institute of Software, CAS, China), Shan Zhao (Institute of Software, CAS, China), and Mingshu Li (Institute of Software, CAS, China)</i>	
RunSAFER: A Novel Runtime Fault Detection Approach for Systolic Array Accelerators	596
<i>Eleonora Vacca (Politecnico di Torino, Dipartimento di Automatica e Informatica, Italy), Giorgio Ajmone (Politecnico di Torino, Dipartimento di Automatica e Informatica, Italy), and Luca Sterpone (Politecnico di Torino, Dipartimento di Automatica e Informatica, Italy)</i>	

Session 12B: Compression & Accelerators

BIRD: A Lightweight and Adaptive Compressor for Communication-Efficient Distributed Learning Using Tensor-Wise Bi-Random Sampling	605
<i>Donglei Wu (Harbin Institute of Technology, Shenzhen), Weihao Yang (Harbin Institute of Technology, Shenzhen), Cai Deng (Harbin Institute of Technology, Shenzhen), Xiangyu Zou (Harbin Institute of Technology, Shenzhen), Shiyi Li (Harbin Institute of Technology, Shenzhen), and Wen Xia (Harbin Institute of Technology, Shenzhen; Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies; Peng Cheng Laboratory, Shenzhen)</i>	
MultiFuse: Efficient Cross Layer Fusion for DNN Accelerators with Multi-Level Memory Hierarchy	614
<i>Chia-Wei Chang (National Tsing Hua University, Taiwan), Jing-Jia Liou (National Tsing Hua University, Taiwan), Chih-Tsun Huang (National Tsing Hua University, Taiwan), Wei-Chung Hsu (National Taiwan University, Taiwan), and Juin-Ming Lu (Industrial Technology Research Institute, Taiwan)</i>	
DEQ: Dynamic Element-Wise Quantization for Efficient Attention Architecture	623
<i>Xuhang Wang (Shanghai Jiao Tong University, China), Zhuoran Song (Shanghai Jiao Tong University, China), Qiyue Huang (Shanghai Jiao Tong University, China), and Xiaoyao Liang (Shanghai Jiao Tong University, China)</i>	
CNN Inference Accelerators with Adjustable Feature Map Compression Ratios	631
<i>Yu-Chih Tsai (National Tsing Hua University, Taiwan), Chung-Yueh Liu (National Tsing Hua University, Taiwan), Chia-Chun Wang (National Tsing Hua University, Taiwan), Tsen-Wei Hsu (National Tsing Hua University, Taiwan), and Ren-Shuo Liu (National Tsing Hua University, Hsinchu, Taiwan)</i>	
Author Index	635